**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 16 June 2025 |
| Team ID | SWTID1749709635 |
| Project Title | Mental Health Prediction |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification Template**

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

**Data Collection Plan Template**

|  |  |
| --- | --- |
| **Section** | **Description** |
| Project Overview | The machine learning project aims to predict whether an individual has sought treatment for mental health issues based on survey responses. Using a dataset with features such as age, gender, country, employment status, family history, work interference, and workplace support, the objective is to build a model that accurately classifies mental health treatment status, supporting awareness and intervention efforts. |
| Data Collection Plan | Search for datasets related to mental health, workplace mental health, and employee well-being.  ● Prioritize datasets with diverse demographic and occupational information.  ● Ensure the dataset includes both personal (age, gender, country) and workplace (remote work, benefits, supervisor support) features. |
| Raw Data Sources Identified | The raw data source for this project is a mental health in tech survey dataset, obtained from Kaggle (and/or similar open data repositories). The provided sample data represents a subset of the collected information, encompassing variables such as age, gender, country, employment status, family history, work interference, and workplace support for machine learning analysis. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Kaggle Dataset | The dataset comprises survey responses from tech employees, including demographic details (age, gender, country), employment status, family history, workplace support, and mental health treatment status. | <https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey> | CSV | 303.68KB | Public |